whole blood cells. Lab. Invest. *79*:1731-1732.
6. **Verhofstede, C., K. Fransen, D. Marissens, R. Verhelst, G. van der Groen, S. Lauwers, G. Zissis, and J. Plum.** 1996. Isolation of HIV-1 RNA from plasma: evaluation of eight different extraction methods. J. Virol. Methods *60*: 155-159.
7. **Chomczynski, P.** 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. BioTechniques *15*:532-537.
8. **Chomczynski, P. and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. *162*:156-159.
9. **Laemmli, U.K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*:8680-8685.
10. **Smallwood, A., A. Papageorghiou, K. Nicolaides, M.K. Alley, J. Alice, G. Nargund, K. Ojha, S. Campbell, and S. Banerjee.** Temporal regulation of the expression of *Syncytin* (*HERV-W*), maternally imprinted *PEG10*, and *SGCE* in human placenta. Biol. Reprod. (In press). Biol. Reprod. 2003 Mar 5 (Epub ahead of print).

Received 17 March 2003; accepted 28 May 2003.

Address correspondence to Subhasis Banerjee, Harris Birthright Research Centre for Fetal Medicine, King's College Hospital Medical School, Denmark Hill, London SE5 9RS, UK. e-mail: dr_sbanerjee@hotmail.com

Selection of optimal internal controls for gene expression profiling of liver disease

Soyoun Kim and Taeuk Kim
LG Chem Ltd./Research Park, Daejeon, Korea

BioTechniques 35:456-460 (September 2003)

Recently developed technologies such as microarray analysis allow researchers to determine the genome-wide patterns of expressed genes. This information provides insight into complex regulatory networks, enables the identification of new or underexplored biological processes, and implicates genes in various disease processes (1). While microarray analysis provides genome-wide information on relative gene expression, real-time reverse transcription-PCR (RT-PCR) provides quantitative information by the simultaneous measurement of gene expression in many different samples, which makes the technique especially suitable for research questions that require the measurement of expression level changes (2). Compared to conventional quantification methods such as Northern blot analysis, RNase protection assay, or competitive RT-PCR, real-time RT-PCR analysis has the advantages of greater speed, higher throughput, and a higher degree of potential automation (3,4). Nevertheless, all strategies for mRNA quantification require accurate, reproducible normalization. For the correct normalization of gene expression analysis, various strategies have been applied, such as counting cells, total RNA quantitation, and rRNA measurement (3). However, internal control genes are most frequently used to normalize mRNA expression in laboratory experiments. The internal control, usually one of the so-called housekeeping genes (5), should not vary between the tissues or cells under investigation or in response to experimental treatment. However, although housekeeping genes are constant in certain cell types, they can vary in other types (6,7), particularly in clinical samples associated with malignant diseases (5). Thus, the selection of proper control genes for clinical patient samples is vital to gene expression analysis.

In the current study, we used liver tissues from normal, liver cirrhosis (LC), and hepatocellular carcinoma (HCC) patients to examine the expression patterns of housekeeping genes. Liver cancer is the third most deadly cancer worldwide and fifth in the number of cases (8), but the molecular mechanisms of hepatocarcinogenesis are not well understood. Therefore, the number of studies probing global gene expression profiles of HCC or preneoplastic chronic liver disease has increased exponentially in recent years (9,10), and the identification of the optimal internal controls is necessary for correct gene expression profiling of liver diseases.

Table 1 describes the 10 common housekeeping genes and gene-specific primers that were used. To compare the expression levels of each housekeeping gene, we used four different groups of liver tissues: normal liver tissues, LC, nontumor LC tissues from an HCC patient, and carcinoma tissues from an HCC patient. The expression level of the 10 internal control genes was determined by real-time RT-PCR analysis, with RNAs extracted from the same amount of 10 different liver tissue samples. Table 2 shows the expression level of each tissue sample. To evaluate the expression stability of each housekeeping gene between liver tissues, we used the geNorm program (3). The geNorm program determines the most stable housekeeping genes from a set of tested genes and calculates the gene expression normalization factor for all tissue samples, based on the geometric mean of a user-defined number of housekeeping genes. The results are ranked and shown in Table 2. From geNorm analysis, the ubiquitin C (*UBC*) gene showed the most stable expression between 10 liver samples, with the hydromethyl-bilane synthase (*HMBS*) gene a close second. In addition, the stepwise exclusion of the least stable gene showed *UBC* and *HMBS* as the most stable genes (data not shown).

Next, we assessed the stability of the *UBC* and *HMBS* genes in 67 additional normal, LC, nontumor LC tissues from an HCC patient, and HCC samples. The stability of these two genes as assessed by geNorm analysis (M value: 3) was 1.02. Because perfect stable expression should result in M values of zero, the geometric mean of the *UBC* and *HMBS* gene results is a very accurate control for liver tissues, as compared to the other commonly used housekeeping genes shown in Table 2. In a study of 67 liver tissues, analysis using the geNorm program showed that the *UBC*

Table 1. Housekeeping Genes Evaluated and Primer Sequences Used in this Study

Gene	Full Name	Forward Primer[a]	Reverse Primer[a]
RPL13A	Ribosomal protein L13a	5′-CCTGGAGGAGAAGAGGAAAGAG-3′	5′-TTGAGGACCTCTGTGTATTTGTC-3′
TUB	β-Tubulin	5′-TTCCAGCTGACCCACTCTCT-3′	5′-ACAGGGCCTCGTTATCAATG-3′
YWHAZ	Tyrosine 3-mono-oxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	5′-ACTTTTGGTACATTGTGGCTTCA-3′	5′-CCGCCAGGACAAACCAGTAT-3′
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	5′-TGCACCACCAACTGCTTAGC-3′	5′-GGCATGGACTGTGGTCATGAG-3′
B2M	β-2-Microglobulin	5′-TGCTGTCTCCATGTTTGATGTATC-3′	5′-TCTCTGCTCCCCACCTCTAAG-3′
HMBS	Hydromethyl-bilane synthase	5′-GGCAATGCGGCTGCAA-3′	5′-GGGTACCCACGCGAATCAC-3′
HPRT1	Hypoxanthine phospho-ribosyl-transferase1	5′-TGACACTGGCAAAACAATGC-3′	5′-GGTCCTTTTCACCAGCAAGCT-3′
ACTB	β-Actin	5′-CTGGAACGGTGAAGGTGAC-3′	5′-AAGGGACTTCCTGTAACAATGC-3′
UBC	Ubiquitin C	5′-CCTGGTGCTCCGTCTTAGAG-3′	5′-TTTCCCAGCAAAGATCAACC-3′
CTBP1	C-Terminal binding protein 1	5′-TTCACCGTCAAGCAGATGAG-3′	5′-GGCTAAAGCTGAAGGGTTCC-3′

[a]Primer sequences were modified from Reference 3.

Table 2. Housekeeping Gene Stability Results by RT-PCR

	Expression Levels of Genes from Same Amount of Liver Tissues[a] (Copy Numbers: 10^3)											
Tissues/ Genes	1 N	2 LC	3 LC	4 LC	5 NT	6 NT	7 NT	8 HCC	9 HCC	10 HCC	geNorm Stability Result[b]	Rank[c]
RPL13A	1095	1950	853	69	973	1142	973	927	15064	4048	2.50	5
TUB	12	23	19	100	9	13	10	58	0.001	6	6.13	10
YWHAZ	67	163	72	137	60	261	70	108	323	84	2.87	8
GAPDH	200	490	228	60	212	210	212	781	880	465	2.43	4
B2M	1016	3067	2334	5	1204	2432	1626	3894	805	445	2.60	6
HMBS	3	7	4	0.09	3	4	4	10	4	8	2.18	2
HPRT1	6	9	4	0.01	4	5	3	3	14	7	2.26	3
ACTB	1294	3305	590	2749	1358	788	671	0.001	4169	1594	6.05	9
UBC	431	870	212	16	333	384	466	789	589	310	2.17	1
CTBP1	3	10	7	0.01	5	6	6	14	15	6	2.69	7

N, normal liver tissues; LC, liver cirrhotic tissues from a cirrhosis patient; NT, nontumor liver cirrhotic tissues from a hepatocellular carcinoma patient; HCC, carcinoma tissues from a hepatocellular carcinoma patient.

[a]The expression levels of each gene were obtained using real-time reverse transcription-PCR (RT-PCR) (see Table 3). Same RNA samples from each tissue were analyzed for 10 different housekeeping gene candidates. Whole samples were analyzed in the same real-time RT-PCR run with standards for each gene. All experiments were performed twice.

[b]For every control gene, the geNorm program determined the pairwise variation, with all other control genes as the standard deviation of the logarithmically transformed expression ratios. M, the internal control gene-stability measure, was defined as the average pairwise variation of a particular gene with all other control genes (3).

[c]Genes with the lowest M values have the most stable expression. Control genes are ranked by their expression stability.

and *HMBS* gene pair provides the most accurate normalization factor for real time RT-PCR analysis, which suggests that these housekeeping genes are the best selection for the future expression profiling of liver tissues.

No single gene was found to vary by less than 2-fold across a panel of 60 cell lines on an 8000 feature array (3), and genes frequently used for normalization, such as *GAPDH* and β-actin (*ACTB*), varied from 7- to 23-fold (6). In particular, tumor tissues showed large variations in housekeeping genes commonly used for normalization (5,7). Taken together, these results suggest that the ideal and universal normalization genes may not exist and indicate that researchers should search for stably expressed genes specific to each experimental system. Because studies in the global gene expression of liver disease have quickly provided rich information and some additional clues to the genesis of liver cancer, this study provides useful information to researchers for selecting a pair of common housekeeping genes for use in malignant liver disease studies.

Table 3. Protocol of Housekeeping Gene Selection for Normalizing Gene Expression in Liver Tissues

1. **Preparation of tissue samples and RNA.** We prepared a total of 67 tissue specimens, including 5 normal liver tissues, 22 liver cirrhotic (LC) tissues from a cirrhosis patient, 20 nontumor LC tissues from a hepatocellular carcinoma (HCC) patient, and 20 carcinoma tissues from an HCC patient. As controls, five normal liver tissues were obtained during partial hepatectomies of patients with other malignant diseases. The samples were ground into a fine powder while still nitrogen-frozen, and total RNA was extracted using the TRIZOL® reagent (Sigma, St. Louis, MO, USA) following the manufacturer's instructions.
2. **Real-time reverse transcription-PCR (RT-PCR) analysis.** Reverse transcription products (cDNA) made with oligo(dT) primers were used as real-time PCR templates with the ThermoScript™ RT-PCR system (Invitrogen, Carlsbad, CA, USA). Real-time PCR analysis was performed as previously described (11). In brief, the templates (cDNAs) from each sample and primer sets from each housekeeping gene were mixed with 2× QuantiTect™ SYBR® Green PCR Master Mix (Qiagen, Valencia, CA, USA), and 40 PCR cycles were performed using a Rotor-Gene™ real-time PCR machine (Corbett Research, Sydney, NSW, Australia). Four standard samples with known copy numbers (10^5, 10^6, 10^7, and 10^8 copies) were used for each gene, and copy numbers were calculated automatically by the Rotor-Gene program.
3. **The geNorm program analysis.** The results from real-time RT-PCR were transformed into the geNorm input worksheet of the geNorm program for gene stability analysis (3), and the output of this program was ranked as described in Table 2. The geometric mean was also calculated to determine the best pair for normalization.

REFERENCES

1. **Schena, M., D. Shalon, R.W. Davis, and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*: 467-470.
2. **Heid, C.A., J. Stevens, K.J. Livak, and P.M. Williams.** 1996. Real time quantitative PCR. Genome Res. *6*:986-994.
3. **Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*:research0034.1-0034.11
4. **Rajeevan, M.S., S.D. Vernon, N. Taysavang, and E.R. Unger.** 2001. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. J. Mol. Diagn. *3*:26-31.
5. **Suzuki, T., P.J. Higgins, and D.R. Crawford.** 2000. Control selection for RNA quantitation. BioTechniques *29*:332-337.
6. **Warrington, J.A., A. Nair, M. Mahadevappa, and M. Tsyganskaya.** 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol. Genomics *2*:143-147.
7. **Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, et al.** 1999. Housekeeping genes as internal standards: use and limits. J. Biotechnol. *75*:291-295.
8. **Parkin, D.M., P. Pisani, and J. Ferlay.** 1999 Global cancer statistics. CA Cancer J. Clin. *49*:33-64.
9. **Kim, J.W. and X.W. Wang.** 2003. Gene expression profiling of preneoplastic liver disease and liver cancer: a new era for improved early detection and treatment of these deadly diseases? Carcinogenesis *24*:363-369.
10. **Chen, X. S.T. Cheung, and S. So.** 2002. Gene expression patterns in human liver cancers. Mol. Biol. Cell *13*:1929-1939.
11. **Kim, S., H. Shi, D.K. Lee, and J.T. Lis.** 2003. Specific SR protein-dependent splicing substrates identified through genomic SELEX. Nucleic Acids Res. *31*:1955-1961.

Received 23 April 2003; accepted 29 May 2003.

Address correspondence to Soyoun Kim, LG Chem Ltd./Research Park, Moonji-dong Dajeon, 305-380, Korea. e-mail: sykim@lgchem.com